Draft Genome Sequence of *Escherichia coli* E1728 Isolated from Marine Sediment in Hong Kong

Jennifer Y. H. Lai,[a] Hao Zhang,[a] Miranda H. Y. Chiang,[a] Min Yu,[a] Rui Zhang,[b] Stanley C. K. Lau[a]

Division of Life Science, the Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong[a]; State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen, Fujian, China[b]

J.Y.H.L. and H.Z. contributed equally to this work.

Recent findings of *Escherichia coli* persisting autochthonously in environmental matrices outside animal bodies have revealed largely unknown facets of the lifestyle and ecophysiology of the species that have yet to be explored. Here, we report the draft genome sequence of *E. coli* E1728 isolated from marine sediment.

Address correspondence to Rui Zhang, ruizhang@xmu.edu.cn, or Stanley C. K. Lau, scklau@ust.hk.

*E*scherichia coli naturally inhabits the lower intestines of warm-blooded animals. Once egested into the external environment, it is faced with a large number of stresses that are different from those inside the animal hosts. Therefore, it was commonly believed that *E. coli* cannot survive for long in the external environment. However, many recent studies have reported *E. coli* populations that persisted autochthonously in soils and sediments without any known association with fecal contamination (1). It was estimated that on the global scale, half of the total population of naturally occurring *E. coli* exists outside animal hosts (2). Therefore, the study of *E. coli* as it thrives in external environments is essential to the understanding of the evolution and lifestyle of the species as a whole.

We isolated *E. coli* E1728 from the sediment of an intertidal mudflat in Hong Kong. Its genomic DNA was extracted using the Wizard genomic DNA purification kit (Promega, USA) and sequenced on an Illumina HiSeq 2000 system using paired-end libraries with 300-bp inserts. The raw reads were *de novo* assembled manually into contigs using the CLC Genomics Workbench version 6.5.1 (CLC bio, USA). A total of 170 contigs were obtained for E1728, ranging from 509 to 364,372 bp each (N_{50}, 131,760 bp), with a total size of 5,264,979 bp and a G+C content of 50.3%. The number of protein-coding sequences (CDSs) predicted using GeneMarkS was 5,135 (3). The total length of the CDSs comprises 88.17% of the total contig size. Homologous comparison of the CDSs using BLAST identified 3,878 genes in 22 COG functional groups and 3,878 genes in 31 KEGG metabolic pathway groups.

Using tRNAscan (4) and RNAmmer (5), E1728 was found to carry 79 tRNAs and 3 rRNAs.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank under the accession no. JDFV00000000. The version described in this paper is the first version, JDFV01000000.

REFERENCES

1. **Ishii S, Sadowsky MJ.** 2008. *Escherichia coli* in the environment: implications for water quality and human health. Microbes Environ. **23**:101–108. http://dx.doi.org/10.1264/jsme2.23.101.
2. **Savageau MA.** 1983. *Escherichia coli* habitats, cell types, and molecular mechanisms of gene control. Am. Nat. **122**:732–744. http://dx.doi.org/10.1086/284168.
3. **Besemer J, Lomsadze A, Borodovsky M.** 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. **29**:2607–2618. http://dx.doi.org/10.1093/nar/29.12.2607.
4. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**:955–964. http://dx.doi.org/10.1093/nar/25.5.0955.
5. **Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–3108. http://dx.doi.org/10.1093/nar/gkm160.